

Mortality Reduction after Oral Azithromycin

Statistical Analysis Plan

UCSF Francis I. Proctor Foundation

The Carter Center

Niger Ministry of Health

Johns Hopkins University Dana Center and Bloomberg School of Public Health

Tanzania National Institute for Medical Research and Muhumbili Medical Center

London School of Hygiene and Tropical Medicine

University of Malawi, Blantyre College of Medicine

Malawi Ministry of Health

Introduction

This document (Statistical Analysis Plan, SAP) describes the planned analysis and reporting for the clinical trial, **Mortality Reduction after Oral Azithromycin**. It includes specifications for the statistical analyses and tables to be prepared for the interim and final Clinical Study Report. This study is a Phase IV clinical trial to compare methods to reduce childhood mortality using mass administration of azithromycin (Pfizer, CAS 83905-01-5) compared to placebo. The content of this Statistical Analysis Plan meets the requirements stated by the US Food and Drug Administration and conforms to the American Statistical Association's Ethical Guidelines.

The following documents were reviewed in preparation of this Statistical Analysis Plan:

- Mortality Reduction after Oral Azithromycin, Protocol
- Mortality Reduction after Oral Azithromycin, Manual of Operations
- ICH Guidance on Statistical Principles for Clinical Trials

The planned analyses described in this SAP will be included in future manuscripts. Exploratory analyses not necessarily identified in this Statistical Analysis Plan may be performed to support the analysis. Unplanned analyses not delineated in this Statistical Analysis Plan will be documented as such in the final Clinical Study Report and manuscripts.

This document will be reviewed prior to the enrollment of patients. All subsequent changes will be indicated by detailed change log in the Appendix.

1 Summary

1.1 Mortality Trial

The trial profile is given in the Manual of Operations, Ch. 3. In the mortality trial, communities are randomized to two arms: **Azithro** and **Control**. The trial is a placebo controlled, double masked, community randomized, multicenter clinical trial.

1.2 Morbidity Trial

The trial profile is given in the Manual of Operations. In the morbidity trials, communities are randomized to two arms: **AzithroPlus**, or **ControlPlus**. At each study site, 15 communities each are randomized to **AzithroPlus** and to **ControlPlus**. The trial is a placebo controlled, double masked, community randomized, multicenter clinical trial. The profile is given in the Manual of Operations, Ch. 3.

2 Statistical Analysis

2.1 Planned Analyses

2.1.1 Mortality Trial

The primary prespecified analysis will use the number of deaths in each community as the outcome variable. For analysis, a death is counted when an individual in the target age range 1–59 months (inclusive) is on census and is not on the next census due to death. Census 0 takes place at the beginning of the study, and each subsequent census takes place approximately six months later. The outcome variable is derived by summing the deaths in each period.

Primary Analysis (Mortality Study).

The primary analysis will be conducted as negative binomial regression. The number of person-years (for persons in the target age range) is derived by summing the person-times for each period. The community-level predictor consists of treatment arm for the community. Specifically, the treatment variable is a community level variable, not an individual level variable. An additional categorical regressor will be included for the country. The likelihood-ratio will be computed, and the statistical significance assessed by permutation testing. Specifically, we will assess the significance of including treatment arm at a two-sided level of 0.05 (a mortality effect in either direction will be statistically significant). The number of permutations is chosen as 10,000, large enough to minimize Monte Carlo error; consecutive runs using a prespecified seed will be reported (and if the Monte Carlo confidence interval for the P-value includes the value 0.05, a new set of 100,000 runs will be conducted and the value reported). An estimated aggregation parameter over 30 will result in the choice of Poisson regression. Evidence of poor fit (based on deviance and/or score residuals) will result in reporting of additional models, beginning with estimation of country-specific aggregation parameters. (Note that the primary analysis remains the single multi-country negative binomial regression, with permutation-based significance testing.)

Individuals who are in the target age range at the beginning of any study period, but who are no longer in the target age range by the end, are considered to be fully enrolled and contribute the full amount of person-time for that period.

All three sites will be aggregated together for the single primary prespecified analysis. The intent is a single negative binomial regression for all three countries together, as the single outcome of the trial.

Model adequacy will always be checked (e.g. overdispersion for count models), and leave-one-out influence plots will be conducted to identify highly influential single observations. All analyses in this trial will always account for the clustered (community randomized) design.

In the event of an unanticipated rare event (such as a natural disaster or bus accident) in part of an area, sensitivity analysis will be conducted. Analysis omitting this region will be reported, as well as a stratified permutation test in which we compare treatment and control within strata defined by the affected region. While entirely unsuitable for primary analysis, under certain circumstances it may be possible to remove mortality due to the rare event; we would propose reporting this analysis as well.

Secondary Analyses (Mortality study)

All secondary analyses will be sharply distinguished from the primary prespecified analysis and are as the following:

1. Age, treatment group, and country will be used as a covariates in the primary analysis model.

2. Negative binomial regression for each time period will be analyzed, using time, country, and treatment.
3. A Poisson regression at the individual level comparing treatment groups, using each period of time.
4. Negative binomial regression contrasting the first and second years.
5. Negative binomial regression using a country-treatment interaction.
6. Treatment effect by age will be analyzed using subgroups of patients defined by the following age brackets: [1, 6), [6-12), [12-18), [18-24), [24, 60). Age of the child will be the age captured at enrollment.
7. We will compare the cause of death, as determined by Verbal Autopsy, between the two arms, using a 2 x k table, permuting by randomization unit.
8. Analyses will be conducted for each country, using data from that country only. This analysis will be negative binomial regression with treatment assignment for each village as the covariate.
9. An on-treatment analysis of mortality will be restricted to those taking the medication (or placebo) (i.e., comparing placebo-takers to medication-takers).
10. Analysis of those censused but too young to treat will be conducted.
11. Data will be formatted to include the Phase 24 vital status of all persons found alive at any of the Phases 0, 6, and 12 (with no indication of them dying in the interim) instead of *only considering the vital status of persons alive at phase 18* census.
12. Subgroup analysis of children eligible for 4 treatments, but receiving only 1, 2, 3, or 4 treatments by arm (1 vs. 1, 2 vs. 2, 3 vs. 3, and 4 vs. 4 treatments).
13. A secondary analysis will be conducted on every child who died and had a treatment (antibiotics or placebo) assessing if:
 - a. The time of death differs between arms (within a year), using a test of equality of circular distributions (using the phase angle as the outcome variable), with a supplementary comparison conducted by fitting negative binomial regression based on linear predictor $A_i \sin(\theta t + \phi)$, where the null hypothesis is that the phase offset ϕ is independent of treatment (similarly, we may compare equality of circular means using a two-sample Edwards test). Note that in this analysis, a test of the equality of the amplitude A_i will also be reported. We will conduct goodness of fit tests using higher order trigonometric terms.
 - b. The time between treatment and death differs between treatment arms will be compared by performing a survival analysis that accounts for country. Note that in order to do a survival analysis, we will include children that survived in the analysis.

- c. The times of treatment (in children that died) between the two arms a Wilcoxon rank sum test comparing the time of treatment in the deceased between the two groups using the Wilcoxon rank sum test.
- 14. Spatial analysis. Secondary analysis will include geographic region as follows. First, we will present a plot of the reduction in mortality by region within country (e.g. zone in Malawi) versus the mortality rate in the placebo communities in that region. This will be done across all countries. Additional data (poverty indices) may be included in such analyses. Such analyses, exploratory in nature, will be sharply distinguished from prespecified outcomes.
- 15. Analyses will be reported using dose (or recorded height/weight as applicable). We will also conduct models in which the number of received treatments is used as a predictor. This is a secondary analysis.
- 16. As stated elsewhere, we will report results of comparing mortality per protocol across arms. We will also report a comparison of people who received treatment (whether placebo or azithromycin) to those who did not. The per-protocol analysis is an important secondary analysis whose results will be highlighted. Comparison of participants to nonparticipants is expected to yield a large effect, but we expect little power for the clinical trial (azithromycin vs placebo) comparison.
- 17. Alternative outcomes will be considered including:
 - a. Verbal autopsy using permanova to test the significance of the treatment variable. Missing verbal autopsy data will not be imputed.
- 18. Subgroup analysis measuring the effect of biannual azithro vs biannual placebo on mortality among subgroups of malnutrition in the MORDOR mortality communities in all countries

2.1.2 Morbidity Trials

Primary Analyses (Morbidity Study).

The morbidity analyses are to be conducted on a country-specific basis. Each outcome will be addressed in a separate analysis. Outcomes are taken at 0, 12, and 24 months.

The outcome and analysis plan for all non-core studies must be submitted to the Steering Committee for approval.

The outcomes for this study are listed on www.clinicaltrials.gov. Analyses may be conducted at the village level (accounting thus for clustering), or, when possible, at the individual level accounting for clustering by using mixed effects regression. The listed primary outcomes (4 February 2015) are:

1. Height over time in children aged 1-60 months (Niger & Malawi)
 2. Weight, corrected for height over time in children aged 1-60 months (Niger & Malawi)
 3. Fraction of macrolide resistance in nasopharyngeal swabs in children 1-60 months (core). Note that for Niger, this is phenotypic, and is a binary outcome at the individual level.
 4. Fraction of macrolide resistance from stool specimens in children 1-60 months (core). For Niger, this includes RNA, DNA, and TAC-card methods (a binary outcome at the individual level).
 5. Fraction of macrolide resistance by nasal swabs in children 1-60 months (Niger only). For Niger, this is based on RNA and DNA sequencing (a binary outcome at the individual level).
 6. Presence of malaria parasites in children 1-60 months (core activity)
 7. Fraction of conjunctival swabs yielding ocular chlamydia in children 1-60 months (Niger)
- Nonnormality, heteroskedasticity, clustering, and outliers will be assessed appropriately.

Secondary Analyses (Morbidity Study).

Common templates will be used. Analysis will use cluster-level variables for malaria parasites, macrolide resistance in NP swabs, macrolide resistance in nasal swabs, macrolide resistance in stool specimens, and ocular chlamydia. Clustered logistic regression using baseline values will be used, with the outcome being the 12 and 24 month observations, and time and treatment as covariates in addition to baseline values. Permutation P-values will be reported.

Model checking will be performed to identify poorly fitting models, using the residuals versus fitted values plot and by finding high leverage points. Significance testing will be conducted using permutation tests.

The analysis will follow these guiding principles: (1) All data from all available participants will be included, regardless of whether the patient contributed data to all time periods, or to only a subset of periods; (2) The analysis must reflect the longitudinal nature of the followup and the clustered nature of the design, (3) Permutation testing will be used to compute the P-value, (4) The model must be fully prespecified prior to any analysis of any kind.

If a village randomized to the morbidity study has fewer than 40 children upon census, an additional morbidity village may be recruited in its place.

All secondary analyses are identified as the following:

1. Repeated measures of outcomes for children present at baseline and possibly followed for 24 months. using 0, 12, and 24-month measurements (or all time points processed).
2. Analyze additional outcomes: WHO z-values (height for age, weight for height, weight for age) and MUAC. P-value will be adjusted for these additional outcomes.
3. Separate outcome using cross-sectional post-test children 25 months and under using 0.05 alpha. Analysis will be similar to longitudinal (height for age, weight for height, outliers, etc. Note that these children were not eligible for the primary analysis.
4. Community-level variables will be considered from CSI records.
5. Per-protocol analysis will be performed by arm, i.e. only including those who took the treatment.
6. Other outcomes by treatment arm at 24 months will include percent wasted and percent stunted. Permutation p-values will be computed here.

Sensitivity analysis including the outliers as well as dissociating the measurements in a repeated measures analysis will be conducted. Appropriate consideration will be analyzed for transformation, heteroskedacity, and outliers. Additional prespecified secondary analyses are listed on www.clinicaltrials.gov.

2.1.3 Microbiome

2.1.3.1 Sample Types

There will be five types of samples collected for microbiome analysis, rectal, nasal-pharangeal, nasal, conjutiva, and blood.

2.1.3.2 Sequencing and Pooling

Samples will be processed using either RNA-sequencing or 16S sequencing. In some instances, specimens will be pooled in the lab by combining aliquots of equal nucleic acid content by QuBit measurement of randomly selected individual's swabs into one microbiome sample per village. There will be one pooled sample of randomly selected specimens (e.g. 10 specimens) for each of the 30 villages, giving us 30 samples total for sequencing.

2.1.3.3 Time Points

Samples for a pre-morbidity study also know internally as “mini-microbiome”, were collected in a separate individually randomized trial. The Morbidity study will provide microbiome samples collected at 12-months and 24-months.

2.1.3.4 Primary Microbiome Analysis

Primary Analysis for rectal swabs collected pre-morbidity study, 16S-sequencing

Pre-Morbidity RCT (a.k.a. “mini-microbiome”) will analyze 80 1-60 month old kids, randomized to placebo or azithromycin. We will obtain pre-treatment rectal swabs on day 0 and

post-treatment rectal swabs on day 5. We will compare diversity between arms at day 5, and between time points in each arm. Specific diversity measures to be examined are Shannon's Index (primary), Simpson's Index.

Primary Analysis for rectal swabs collected at 12-months, RNA-sequencing

Primary Analyses for the 12 month rectal swabs will be conducted on pooled samples. We will use PERMANOVA on an **L2-norm** distance on bacterial **reads** to determine if treatment significantly explains the difference between the two groups.

Primary Analysis for rectal swabs collected at baseline and at 24 months, RNA and DNA sequencing; individual level analysis, not pooled. We propose to conduct PERMANOVA based on L2-norm; the outcome is the distance in centroid of the whole sequences, comparing the two treatment groups. Primary analysis is posttest only, but we will report comparisons at baseline as well.

Primary Analysis for conjunctiva swabs, RNA-sequencing

Primary Analyses for the 24 month **conjunctiva swabs-RNA sequencing** will be conducted on pooled samples. We will use PERMANOVA on an **L2-norm** distance on bacterial **reads** to determine if treatment significantly explains the difference between the two groups.

Primary analysis for rectal swabs at 36-months, DNA-sequencing

Primary analysis on the rectal swabs will be conducted on pooled samples. We will be using DESeq2 to determine if *Campylobacter upsaliensis* is upregulated in the placebo arm compared to the azithromycin arm. LEfSe will be used as a sensitivity test.

2.1.3.5 Secondary Microbiome Analysis

Secondary Analysis for all types and timepoints of microbiome samples will include:

- L1-norm distance PERMANOVA on bacterial **reads**
- For RNA-sequenced data: L1, L2-norm distances on virus, parasites, fungi.
- Simpson's and Shannon's estimated gamma diversities (transformed into effective number) will be compared by arm using a permutation test without correcting for baseline (ie post test only)
- Conjunctiva swabs, 16S-sequencing at 24-months will be conducted on individual samples. We will use PERMANOVA on an L1-norm distance on bacterial OTUs to determine if treatment significantly explains the difference between the two groups.
- TAC-card for rectal specimens of bacteria and viruses between two treatment groups at 24 months and at baseline (PERMANOVA), and similarly for NP swabs (i.e., TAC-card)

Secondary analysis for rectal swabs at 36-months

We propose to compare the normalized abundance of MLS reads between treatment arms. Resistance to other classes of antibiotics will be compared. In addition, we will also assess for a

difference in the normalized abundance of betalactams reads. P-values will NOT be corrected for multiple comparisons. Instead, we will estimate the FDR for resistance to each class of antibiotic.

2.1.3.6 Clinic-Based Case Findings

Clinic-based case finding will use consecutive cases from all CSIs in Loga district. Consecutive cases of infectious illnesses among 0-5 year-old children will be assessed from the case log from months 0 to 12. The number of cases of infectious illness per randomization unit will be compared for 1 calendar year between the intervention and control groups in a Poisson regression, adjusting for the number of cases of infectious illness recorded in the calendar year preceding the start of the intervention.

2.1.3.7 Clinic-Based NP swabs

Clinic-based NP swabs. We will collect nasopharyngeal swabs from children residing in morbidity communities who present to the CSI with respiratory complaints within the month of the annual morbidity assessment (i.e., month 0, 12, 24), and process swabs for macrolide resistance. The primary analysis will compare macrolide resistance in the azithromycin versus placebo groups using mixed effects regression as described above. As a secondary analysis, we will compare the prevalence of macrolide resistance in clinic-based swabs vs. community-collected swabs in a mixed effects regression, with a permuted p-value that accounts for the stratified nature of the data.

2.3 General considerations

Individual level missing data (due to loss to follow-up or dropouts) is expected. If an individual is lost to follow-up or otherwise unavailable during an interval of length D , then the individual is assumed to contribute the amount of person-time $D/2$ to the denominator (midpoint imputation).

Village-level missing data could occur for various reasons, including field error, withdrawal of support from village leadership, and so forth. While this has not happened in previous studies in our group, the analytic plan will be to analyze all available information (including data from any village prior to the time of dropout of that village). Sensitivity analysis will be conducted in which various values are imputed to the missing observations to determine how our conclusions would change.

3 Sample Size Planning

3.1 Mortality trial

We use the following formula (cited from Cluster Randomised Trials Hayes RJ, Moulton LH),:

$$c=1+ \left(z_{\alpha/2}+z_{\beta} \right)^2 \frac{(\lambda_0+\lambda_1)/y+k^2(\lambda_0^2+\lambda_1^2)}{(\lambda_0-\lambda_1)^2}$$

where

- c is the number of clusters per arm,
- y is the number of person-years per cluster,
- λ_0 is the rate in the control group,
- λ_1 is the rate in the treatment group,
- k is the coefficient of variation in both groups,
- α is the alpha level,
- β and is the type II error rate.

3.1.1 Malawi

- The base rate for mortality is assumed to be 0.0148 per year in the target age range.
- We assume the same coefficient of variation for village-level mortality as for the Niger site, i.e. 0.51.
- We assume that 19% of the population is in the target range of 1–59 months (inclusive) for the Malawi site.
- We assume an average community size of 799 people.
- We assume 10% loss to follow-up per year.
- The effect size is 0.15.
- The desired power is 80%.

The number of clusters per arm for Malawi is 310, the total number of clusters for Malawi is 620, and the expected total number of children involved is 94122. In practice, we allow that these are estimations and will depend on drug/placebo availability, with the request being for 83,333 doses per country, assuming 3 doses per 30ml bottle of suspension.

Supplementary assessment, June 2016: In practice the Malawi villages effective size is 1464 individuals, with a slightly higher baseline mortality rate of 0.005 per year. The sample size of 306 randomization units in both arms appears to give approximately 41% power for an effect size of 15% for Malawi alone.

3.1.2 Niger

The input assumptions are as follows:

- The base rate for mortality is assumed to be 0.02 per year in the target age range, lower than seen in the PRET study in Niger.
- We assume the coefficient of variation for village-level mortality is 0.51 (derived from PRET Niger data).
- We assume that 16.7% of the population is in the target range of 1–59 months (inclusive) for the Niger site. National statistics suggest a slightly higher fraction (19%), which would yield fewer communities than we are planning for.
- We assume an average community size of 668 people. The communities are known as *grappes*.
- We assume 10% loss to follow-up per year (source: PRET Niger, Z. Zhou, pers. comm.).
- The effect size is 0.15.
- The desired power is 80%.

The number of clusters per arm for Niger was projected to be 312, the total number of clusters for Niger was projected to be 624, and the expected total number of children involved was projected to be 69472. In practice, we allow that these are estimations and will depend on drug/placebo availability, with the request being for 83,333 doses per country, assuming 3 doses per 30ml bottle of suspension. These may be underestimates because of the short interval and the lack of a rainy season.

The final randomization list contained 646 communities, 30 of which were randomized to the morbidity study.

Supplementary assessment, June 2016: In practice the Niger villages appear to have an average of 688 individuals, with a slightly higher baseline mortality rate of 0.025 per year. The sample size of 622 randomization units in both arms appears to give approximately 84% power for an effect size of 15% for Niger alone.

3.1.3 Tanzania

The input assumptions are as follows:

- Considerable uncertainty exists in the base rate for mortality, with national estimates substantially lower than estimates for nearby regions within the country. For planning, we chose a rate of 0.016 per year in the target age range, but recognize that if the true mortality rate is lower, we will have lower power.
- We assume a smaller coefficient of variation for village-level mortality than for the Niger site. Specifically, this value is 0.4, as determined by consultation with the Johns Hopkins team.
- We assume that 17% of the population is in the target range of 1–59 months (inclusive) for the Tanzania site.
- We assume an average community size of 600 people. The communities are known as *hamlets*.
- We assume 10% loss to follow-up per year.
- The effect size is 0.15.
- The desired power is 80%.

The number of clusters per arm for Tanzania was projected to be 307, the total number of clusters for Tanzania was projected to be 614, and the expected total number of children involved was projected to be 62628. In practice, we allow that these are estimations and will depend on drug/placebo availability, with the request being for 83,333 doses per country, assuming 3 doses per 30ml bottle of suspension.

The final randomization list contained 644 communities, of which 30 were randomized to the morbidity study.

Supplementary assessment, June 2016: In practice the Tanzania randomization units appear to have an effective size of individuals, with a slightly higher baseline mortality rate of 0.006 per year. The sample size of 631 randomization units in both arms appears to give approximately 31% power for an effect size of 15% for Tanzania alone.

Note that the primary analysis is all countries together. We anticipate more than 80% power to detect an effect size of approximately 0.10, or ten per-cent.

Update: 2016. We anticipate approximately 80% power to detect an overall effect of 14%, accounting for changes in recruitment and mortality.

3.2 Morbidity Trial

Assuming approximately 12% baseline resistance (based on previous studies) and an ICC of approximately 0.051 (derived from the TEF study), we anticipate approximately 80% power to detect a difference of 18% (i.e., comparing 12% to 30%) assuming 80% carriage of pneumococcus and 10 specimens per community. For resistome comparisons, we anticipate approximately 80% power to detect a 16% difference (comparing 12% to 28%). We anticipate five such comparisons: (1) nasopharyngeal DNA, (2) nasopharyngeal RNA, (3) rectal DNA sequencing, (4) rectal RNA sequencing, and (5) rectal TAC-card.

4 Randomization

Simple random sampling will be used in each country. Within each country, a full list of eligible communities will be provided, and used to fill a column of a table. A second column will be populated with treatment arm names as follows. First, fifteen rows will be filled with the name **AzithroPlus**, and fifteen with the name **ControlPlus**. Then, one half of the remaining rows (rounding to even) will be filled with the name **Azithro**, and the remaining filled with the name **Control**. This column will be shuffled at random, yielding an assignment for each community.

Note that there is no stratification other than country, and the morbidity communities are randomized from the same pool as the mortality communities.

The randomization will be conducted using R. The function `sample` with option `replace=FALSE` will be used to conduct the random shuffling. Note that the choice of the random number seed completely determines the randomization. To ensure the integrity of the randomization, we will choose the seed as follows. We will use the units (least significant digit) of the daily high Fahrenheit temperature in eleven selected US cities on 15 July 2013 as reported by the National Weather Service. A confidential ordered list of these cities will be provided to the DSMC.

Masking will be strictly enforced. Each study site will provide one individual to whom the randomization list will be provided. Investigators and field teams must remain masked and must make no effort to determine which arm a community has been randomized to within the mortality trial or within the morbidity trial.

5 Abbreviations

ANCOVA Analysis of covariance

DSMC Data and Safety Monitoring Committee

HAZ height for age Z score

MUAC Mid upper arm circumference

SAP Statistical Analysis Plan

TF Trachomatous inflammation—follicular

TI Trachomatous inflammation—intense

WAZ weight for age Z score

WHZ weight for height Z score

6 Revisions History

1. **3 March 2016.** Clarified the planned interim analysis for efficacy.
2. **26 April 2016.** Clarification of timing of verbal autopsy collection.
3. **28 June 2016.** Clarification of original power calculations, plus supplementary assessment now that trial is underway.
4. **1 Feb 2017.** Clarified and added secondary analyses for Mortality (page 5).
5. **1 Feb 2017.** Added secondary analyses for Morbidity (page 6).
6. **1 Feb 2017.** Added statistical information for the microbiome analyses for Morbidity (page 7).
7. **6 April 2017.** Page 5, Section 2.1.1. Added the secondary analysis after the DSMC meeting due to the fact that some persons may not be captured at all censuses, but still be alive or dead at the end of phase 24. :

*A secondary analysis will be conducted by doing the following. Data will be formatted to include the Phase 24 vital status of all persons found alive at any of the Phases 0, 6, and 12 (with no indication of them dying in the interim) instead of **only considering the vital status of persons alive at phase 18 census***

8. **24 May 2017 Page 7, Section 2.1.3 Microbiome** Added the primary analysis for the 12-month MB samples.

Primary Analyses for the 12 month rectal swabs will be conducted on pooled samples. Samples will be pooled in the lab by combining equal aliquots of 10 individual's rectal swabs into one microbiome sample per village. There will be one pooled sample of 10 specimens for each of the 30 villages, giving us 30 samples total. We will test for Alpha diversity of the sample (or Gamma diversity of the village) comparing the two arms using a permutation test and not correcting for baseline (post test only).

9. 5 June 2017 Page 7, Section 2.1.3 Microbiome (revised).

Primary Analyses for the 12 month rectal swabs will be conducted on pooled samples. Samples will be pooled in the lab by combining aliquots of equal nucleic acid content by QuBit measurement of 10 individual's rectal swabs into one microbiome sample per village. There will be one pooled sample of 10 specimens for each of the 30 villages, giving us 30 samples total. We will use PERMANOVA on an L1-norm distance on bacterial OTUs (which have undergone rarefaction) to determine if treatment significantly explains the difference between the two groups.

Secondary Analysis will include:

- *L2-norm distance PERMANOVA on bacterial OTUs*
- *virus, parasites, fungi, etc*
- *Simpson's and Shannon's estimated gamma diversities (transformed into effective number) will be compared by arm using a permutation test without correcting for baseline (ie post test only).*

10. 10 June 2017 Page 7, Section 2.1.3 Microbiome (revised).

Analysis for rectal swabs, RNA-sequencing

Primary Analyses for the 12 month rectal swabs will be conducted on pooled samples. Samples will be pooled in the lab by combining aliquots of equal nucleic acid content by QuBit measurement of 10 individual's rectal swabs into one microbiome sample per village. There will be one pooled sample of 10 specimens for each of the 30 villages, giving us 30 samples total for the RNA-seq processing. We will use PERMANOVA on an L2-norm distance on bacterial reads to determine if treatment significantly explains the difference between the two groups.

Secondary Analysis will include:

- ***L1-norm distance PERMANOVA on bacterial reads***
- *virus, parasites, fungi, etc*
- *Simpson's and Shannon's estimated gamma diversities (transformed into effective number) will be compared by arm using a permutation test without correcting for baseline (ie post test only).*

Analysis for conjunctiva swabs, RNA-sequencing

Primary Analyses for the 12 month conjunctiva swabs-RNA sequencing will be conducted on pooled samples. Samples will be pooled in the lab by combining aliquots of equal nucleic acid content by QuBit measurement of 10 individual's conjunctival swabs into one microbiome sample per village. There will be one pooled sample of 10 specimens for each of the 30 villages, giving us 30 samples total for RNA sequencing. We will use PERMANOVA on an L2-norm distance on bacterial reads to determine if treatment significantly explains the difference between the two groups.

Secondary Analysis will include:

- ***L1-norm distance PERMANOVA on bacterial reads***

- *virus, parasites, fungi, etc*
- *Simpson's and Shannon's estimated gamma diversities (transformed into effective number) will be compared by arm using a permutation test without correcting for baseline (ie post test only).*

Analysis for conjunctiva swabs, 16S-sequencing

Primary Analyses for the 12 month conjunctiva swabs-16 sequencing will be conducted on individual samples. We will use PERMANOVA on an L1-norm distance on bacterial OTUs to determine if treatment significantly explains the difference between the two groups.

Secondary Analysis will include:

- *L2-norm distance PERMANOVA on bacterial OTUs*
- *Simpson's and Shannon's estimated gamma diversities (transformed into effective number) will be compared by arm using a permutation test without correcting for baseline (ie post test only).*

11. 29 June 2017 Pages: 4-6 Secondary Analysis for Mortality study were numerated, and clarified. Secondary Analysis for Morbidity study were also numerated, and clarified. Microbiome Section was divided into 4 sections and modified.

12. 3 March 2018 Page 6: Secondary analysis item 15 added per BMGF request (subgroups among malnutrition).

13. 11 June 2018. Sample size calculations from original BMFG proposal updated and transferred to SAP, page 7. The exact analyses for the Niger study have been clarified.

14. 21 August 2018. Additional unplanned secondary analyses for geography and dose added; reporting for per-protocol analysis clarified.

15. 22 April 2019. Page 9, Section 2.1.3 Microbiome (revised).

Primary analysis for rectal swabs at 36-months, DNA-sequencing

Primary analysis on the rectal swabs will be conducted on pooled samples. We will be using DESeq2 to determine if Campylobacter upsaliensis is upregulated in the placebo arm compared to the azithromycin arm. LEfSe will be used as a sensitivity test.

Secondary analysis for rectal swabs at 36-months

We propose to compare the normalized abundance of MLS reads between treatment arms. Resistance to other classes of antibiotics will be compared. In addition, we will also assess for a difference in the normalized abundance of betalactams reads. P-values will NOT be corrected for multiple comparisons.

Changed conjunctival samples sequencing from 12 months to 24 months.

16. 15 July 2019. Page 9, Section 2.1.3.5 Secondary MicrobiomeAnalysis (revised).

Secondary analysis for rectal swabs at 36-months

We propose to compare the normalized abundance of MLS reads between treatment arms. Resistance to other classes of antibiotics will be compared. In addition, we will also assess for a difference in the normalized abundance of betalactams reads. P-values will NOT be corrected for multiple comparisons. Instead, we will estimate the FDR for resistance to each class of antibiotic.

Mortality Reduction after Oral Azithromycin

Contingency Study, Niger, Years 3-4

Statistical Analysis Plan

UCSF Francis I. Proctor Foundation

The Carter Center

Niger Ministry of Health

Johns Hopkins University Dana Center and Bloomberg School of Public Health

Tanzania National Institute for Medical Research and Muhumbili Medical Center

London School of Hygiene and Tropical Medicine

University of Malawi, Blantyre College of Medicine

Malawi Ministry of Health

Introduction

This document (Statistical Analysis Plan, SAP) describes the planned analysis and reporting for the clinical trial, **Mortality Reduction after Oral Azithromycin**, Contingency Study, Years 3-4, Niger. It includes specifications for the statistical analyses and tables to be prepared for the interim and final Clinical Study Report. This study is a Phase IV clinical trial to compare methods to reduce childhood mortality using mass administration of azithromycin (Pfizer, CAS 83905-01-5) compared to placebo. The content of this Statistical Analysis Plan meets the requirements stated by the US Food and Drug Administration and conforms to the American Statistical Association's Ethical Guidelines.

The following documents were reviewed in preparation of this Statistical Analysis Plan:

- Mortality Reduction after Oral Azithromycin, Protocol
- Mortality Reduction after Oral Azithromycin, Statistical Analysis Plan
- Mortality Reduction after Oral Azithromycin, Manual of Operations
- ICH Guidance on Statistical Principles for Clinical Trials

The planned analyses described in this SAP will be included in future manuscripts. Exploratory analyses not necessarily identified in this Statistical Analysis Plan may be performed to support the analysis. Unplanned analyses not delineated in this Statistical Analysis Plan will be documented as such in the final Clinical Study Report and manuscripts.

This document will be reviewed prior to the enrollment of patients. All subsequent changes will be indicated by detailed change log in the Appendix.

MORDOR Stage 2 Contingency Study

1 Introduction

Should the primary analysis of azithromycin to prevent mortality reveal evidence of a benefit, we propose to proceed with two additional studies in Stage 2.

2 Design

Contingency study A. All communities in Niger will receive treatment at the conclusion of MORDOR Stage 1 (original study). Specifically, azithromycin communities will continue with azithromycin, and placebo communities will receive azithromycin at the 24 month visit (census number 5). Follow-up will occur at month 30 and month 36, providing one additional year of data.

Contingency study B. At the 36 month visit, rerandomization will occur, and half of communities will begin to receive placebo.

3 Analysis

The primary outcome is mortality, analyzed in the same way as in MORDOR Stage I, i.e., present on an initial census, and absent on the next follow-up census due to death.

Study A:

1. Sample size considerations.
 - a. Assuming communities of size 668 with approximately 17% in the target age range, an effect size of 20%, 10% loss to follow-up per year, 2% mortality per year, and a CV of 0.51, 312 communities per arm provides approximately 87% power (see MORDOR Phase I SAP for details).
 - b. Updating based on data from Phase I (Niger), we find the following. Assuming the observed CV of 0.34, a death rate of 2% per year results in approximately 80% power to detect a 17% effect. A death rate of 2.5% per year (and a CV of 0.34) result in approximately 80% power to detect a 15.5% effect size.
2. Primary prespecified analysis: We propose to compare the former placebo arms with the former azithromycin arms, using negative binomial regression. Significance testing is two-sided with an alpha of 0.05, and will be based on permutation testing. (This analysis is identical in every way to the country subgroup analyses conducted in Stage 1). This is restricted to the period from 24 months to 36 months, covering two six-month study intervals.
3. We propose, as a prespecified secondary analysis, a longitudinal comparison of the original placebo communities. We propose to contrast the first four study intervals during which these communities received placebo, to the last two study intervals

- during which these communities received azithromycin. Analysis will be conducted using negative binomial regression, with a permutation P-value reported (two-sided, $\alpha=0.05$). The community-randomized nature of the trial is thus taken into account.
4. Additional supplementary analyses will be distinguished from the two prespecified analyses above. In particular, the following will be reported:
 - a. Analysis of the seasonality and timing of deaths: time of death, time of treatment for those who died, and time since treatment for all deaths. We will conduct comparisons of time since treatment conditional on death based on the Cramer-von Mises test (exactly as in Stage 1), additionally comparing across treatment arms and longitudinally. Comparisons of equality of circular means will be used to assess seasonality of death and of treatment, comparing between randomization arms and longitudinally. We will, in particular, focus special attention and reporting to a comparison of phases 1-4 to 5-6 within the placebo arm.
 - b. Analysis will be reported according to the same age classes used in the original study (Keenan et al, NEJM, 2018), i.e. 1-5, 6-11, 12-23, 24-59 months.
 - c. We will examine the difference between phase 5 and 6 in the same way.
 - d. We will report longitudinal analysis of mortality changes in the original azithromycin arm, using a linear predictor of study interval (1–6).
 - e. We will compare estimated treatment effect at a regional level using placebo mortality rates as a predictor.

Study B (year two of mortality contingency study)

1. Sample size considerations are the same as in Study A (year 1).
 - a. Assuming communities of size 668 with approximately 17% in the target age range, an effect size of 20%, 10% loss to follow-up per year, 2% mortality per year, and a CV of 0.51, 312 communities per arm provides approximately 87% power (see MORDOR Phase I SAP for details).
 - b. Updating based on data from Phase I (Niger), we find the following. Assuming the observed CV of 0.34, a death rate of 2% per year results in approximately 80% power to detect a 17% effect. A death rate of 2.5% per year (and a CV of 0.34) result in approximately 80% power to detect a 15.5% effect size.
2. The primary prespecified analysis will contrast the rerandomized placebo arm with the rerandomized azithromycin arm. This will be conducted with negative binomial regression, using a two-sided α of 0.05 and a permutation test by randomization arm.
3. Supplementary analyses will control for prior history (original randomization arm).
4. Longitudinal analysis for communities who always received azithromycin will be reported.
5. We will compare all four groups for the last year of data (two histories by two final treatments).
6. Additional analyses by the same age groups as before and by phase will be reported.
7. Models of timing, seasonality, and geography will be reported as secondary analysis.

Morbidity Study. Analysis of morbidity communities will follow the exact template as Stage I. Primary prespecified analyses will be two-sided, at an alpha of 0.05. However, the morbidity communities are still receiving placebo. Note that we will report the full follow-up at 48 months separately from the Stage 1 primary analyses. We will also look for secular effects longitudinally over the four years of the full study, as well as phases 5-8 separately.

1. Anthropometry: we will examine height and weight longitudinally, and WFH Z score secondarily. Please see the Stage 1 SAP for details.
2. Other analyses will include drug resistance and malaria, as in Stage 1.
3. Note that we do not do nares or conjunctiva swabs, conjunctival photos, or any swabs on 7-12 year olds in the contingency study.

Statistical considerations

1. Communities which drop out of the study will be omitted from analysis for all study period for which we have no data for that community. If a community rejoins at a particular phase, it will be included as before.
2. Evidence of poor model fit to the negative binomial regression will justify reporting additional models.
3. Permutation tests will be conducted with 10000 replications. If the critical value 0.05 lies within the Monte Carlo confidence interval for the P-value, we will replicate with 10,000,000 replications, and this will be reported.